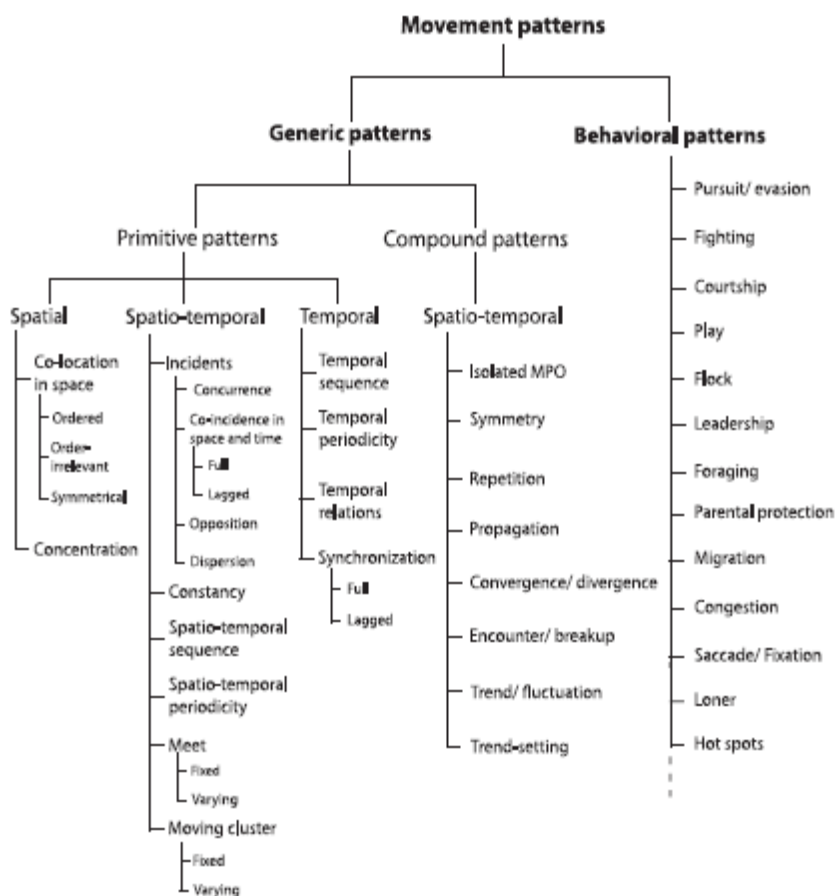


本周主要工作围绕“城市中人群运动模式查询”问题展开。运动对象的研究主要是以 Movement Point Objects 为研究对象（包括车辆、人、动物、自然现象等），研究包括运动轨迹或路线、运动是否受约束、个体运动还是群体运动、运动的不确定性以及运动的时空模式等。对于运动的模式，论文“Towards a taxonomy of movement patterns”中总结如下：

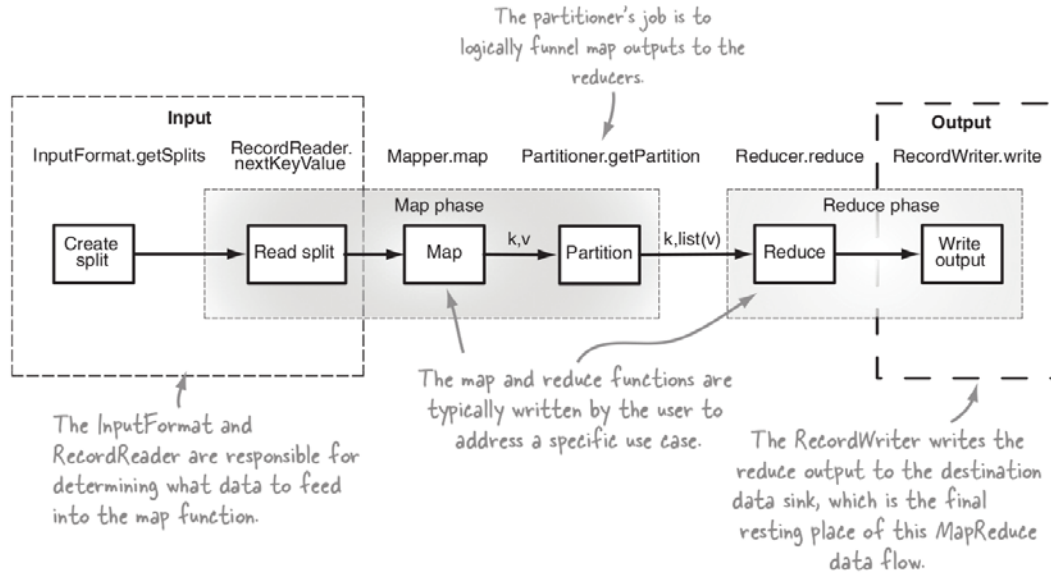


其中 Laube 已经对常规的运动模式如 flock、leadership、Encounter 等有较完整的定义和检测算法。我们的问题依然集中在两个方面：

- 1、数据规模较大的情况下，如何提高检测算法
- 2、人群运动与其他 MPO 有所不同，其模型是否可以有所创新

对于第一个问题，主要是想在 HDFS 上构建索引提高检测和查询效率。本周花费了将近一天多的时间理清了 HDFS 的工作原理，特别是文件分割和读写过程。对于大文件，HDFS 提供了专门的分割接口，将大文件分割后分成多个 Mapper 进行读取，对于特定的

文件格式需要自己定义分割方法（如图 1）。



而文件的写入稍微容易些，但是其中需要注意 **namenode** 和 **datanode** 对于文件的处理上。在文件的写入过程中 **namenode** 其实监控整个写入过程，并保持该文件的元数据，而且这个过程由 **HDFS** 自动完成，如图 2 所示。在掌握这两个原理后，终于修改了先前的代码能够成功运行大文件的分割与写入，为在 **HDFS** 上构建索引扫清了障碍。

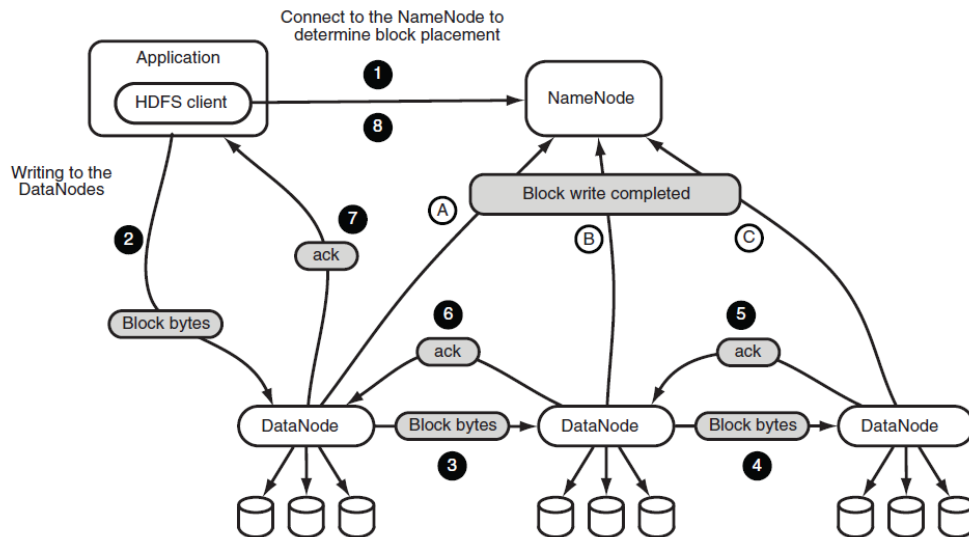


Figure C.2 HDFS write data flow

对于第二个问题，缅因州立大学的 Kathleen Stewart 教授提出，现有的运动模型大都假设时间是一维的，其模型大都是线性的，如描述运动的开始、结束、停顿、移动等，或者

运动中先后发生的事件。然而如果研究未来时间可能发生的运动，那么还需要考虑分叉和合拢的情况。为此，她假设了一种非线性带有分叉和环路的模型，如下图所示。查看其主页发现该项工作只有一篇论文。不过这种分叉、合拢的模型可以用于研究城市人群的移动模式。下周准备仔细研究下她论文的细节,” Modeling alternative sequences of events in dynamic geographic domains”。

